Securing and Supporting Research Projects: Facilitation Design Patterns

NSF Sponsored Workshop - Educause SPC
Date/Time: Monday, May 13, 2019 from 1:00 – 4:30 p.m.

The ResearchSOC is supported by the National Science Foundation under Grant 1840034. The views expressed do not necessarily reflect the views of the National Science Foundation or any other organization.

Introductions, Logistics and Workshop Purpose

- Prominently place color coded dot on your collar
  - GOLD if you're a CISO or other executive
  - RED if you're an engineer or analyst of any sort
  - BLUE if you're an academic / scientist

- Cyd Burrows-Schilling, UCSD, Research Facilitator
- Vlad Grigorescu, ESnet, Security Engineer
- Todd Stone, IU RSOC
- Michael Corn, UCSD, CISO
- Claire Mizumoto, UCSD Research Facilitator
- Tanya Berger-Wolf, Prof. UIC
- Florence D. Hudson, Founder & CEO at FDHint, LLC
- Mary Conley, Trusted CI

Very quick whip around introductions, 5-10 seconds per person. Name, School, Title, and reason for attending.

Ground Rules

- Interrupt any and every time you have a question or comment
  - This is a workshop, not a lecture, your participation is necessary
  - Feel free to disagree!
  - PLEASE share your stories of success or failure
- We'll have one scheduled break at 2:30 for refreshments, but if we're on schedule we'll try to squeeze in a couple of 5 minute breaks
- Slides are or will be available
- Don't forget the post workshop Educause appraisal
- Please feel free to reach out to Cyd or Mike directly with suggestions: this is a 4 hour version of a 3 day version

Workshop Overview

Module 1: Framing and Context

Module 2: Understanding the Researcher and Research Program

Module 3: Research and the Security World

Framing and Context

Group Activity

Research vs. Enterprise Security

Data and Situational complexity

Enterprise vs. Research Security

- At your tables, group together:
- Describe (in columns or a couple of sentences) how Research programs differ from Enterprise/administrative activities
- Report out (choose one person from each group to report out)

Enterprise vs. Research

- More control
- Mostly COTS
- Offices, Data Centers, SaaS
- Mostly staff
- Office culture
- Predictable

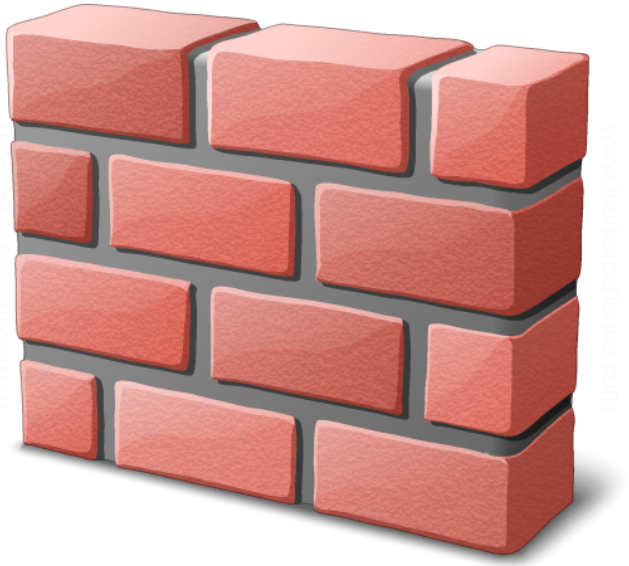
- Less control
- Custom hardware/software lots of macgyvering
- Planes, Trains, and Automobiles
- Faculty
- Lab culture
- Unpredictable (but urgent)

An Example: The Sally Ride

- 40+ vlans
- 3 primary networks (ship, people, instruments)
- At sea ~ 275 days/year
- Cocktail straw satellite connection
- “Day mariners” hired in foreign ports
- $40,000/day to operate
- Many DoD sponsored projects, CUI data
- Connected to port by 10Gbps

We must develop Agility, Flexibility, and understand Data Complexity

CISSP model runs smack into the research program brick wall



- Write policy -> inventory environment -> impose controls

SLOW

If we can’t use control frameworks, are we really just talking about compensating controls?

Right, but even the term “compensating control” is a term of art for security professionals, not researchers

Rigid!

Eg, data classifications

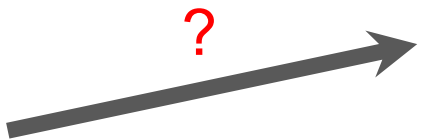
Protection Level classifications:

Protection Level Classification	
Level	**Impact of disclosure or compromise**
P4 - High	Institutional Information and related IT Resources whose unauthorized disclosure or modification could result in significant fines, penalties, regulatory action, or civil or criminal violations. Statutory, regulatory and contract obligations are major drivers for this risk level. Other drivers include, but are not limited to, the risk of significant harm or impairment to UC students, patients, research subjects, employees, guests/program participants, UC reputation, the overall operation of the Location or essential services. (Statutory.)
P3 - Moderate	Institutional Information and related IT Resources whose unauthorized disclosure or modification could result in small to moderate fines, penalties or civil actions. Institutional Information of which unauthorized use, access, disclosure, acquisition, modification, loss or deletion could result in moderate damage to UC, its students, patients, research subjects, employees, community and/or reputation; could have a moderate impact on the privacy of a group; could result in moderate financial loss; or could require legal action. This classification level also includes lower risk items that, when combined, represent increased risk. (Proprietary.)
P2 - Low	Institutional Information and related IT Resources that may not be specifically protected by statute, regulations or other contractual obligations or mandates, but are generally not intended for public use or access. In addition, information of which unauthorized use, access, disclosure, acquisition, modification or loss could result in minor damage or small financial loss, or cause minor impact on the privacy of an individual or group. (Internal.)
P1 - Minimal	Public information or information intended to be readily obtainable by the public, but whose integrity is important and for which unauthorized modification is the primary protection concern. IT Resources for which the application of minimum security requirements is sufficient. (Public.)

- Confidentiality w/out PII
- P-level => specific controls called out in standards
- Focused primarily on institution or individuals as stakeholders, but **where's the science?**

Security for Research: a new triad

- Efficient
  - Rapid response to exigent needs
  - Short-lifetime
  - Collaborative
- Trusted
  - Integrity
  - Defensible
  - Quality Assurance
- Reproducible
  - Repeatable
  - Replicable
  - Reproducible



Protection Level classifications:

Protection Level Classification	
Level	**Impact of disclosure or compromise**
P4 - High	Institutional Information and related IT Resources whose unauthorized disclosure or modification could result in significant fines, penalties, regulatory action, or civil or criminal violations. Statutory, regulatory and contract obligations are major drivers for this risk level. Other drivers include, but are not limited to, the risk of significant harm or impairment to UC students, patients, research subjects, employees, guests/program participants, UC reputation, the overall operation of the Location or essential services. (Statutory.)
P3 - Moderate	Institutional Information and related IT Resources whose unauthorized disclosure or modification could result in small to moderate fines, penalties or civil actions. Institutional Information of which unauthorized use, access, disclosure, acquisition, modification, loss or deletion could result in moderate damage to UC, its students, patients, research subjects, employees, community and/or reputation; could have a moderate impact on the privacy of a group; could result in moderate financial loss; or could require legal action. This classification level also includes lower risk items that, when combined, represent increased risk. (Proprietary.)
P2 - Low	Institutional Information and related IT Resources that may not be specifically protected by statute, regulations or other contractual obligations or mandates, but are generally not intended for public use or access. In addition, information of which unauthorized use, access, disclosure, acquisition, modification or loss could result in minor damage or small financial loss, or cause minor impact on the privacy of an individual or group. (Internal.)
P1 - Minimal	Public information or information intended to be readily obtainable by the public, but whose integrity is important and for which unauthorized modification is the primary protection concern. IT Resources for which the application of minimum security requirements is sufficient. (Public.)

To Von Welch https://bit.ly/2UPVEfc

“Security” must adapt to the structure of science

“The **Sunk Cost Fallacy**. The Misconception: You make rational decisions based on the future value of objects, investments and experiences. The Truth: Your decisions are tainted by the emotional investments you accumulate, and the more you invest in something the harder it becomes to abandon it.” (https://youarenotsosmart.com/2011/03/25/the-sunk-cost-fallacy/)

Understanding the Researcher and Research Program

Intro to Sponsored Projects

Facilitation Skills: use cases and case study

An Introduction to Sponsored Projects

First…. the Faculty

Then… the Funding Landscape

Acknowledgements: Greg Monaco, Kate Adams, Henry Neeman, Dana Brunson, Marcus Bond - content shared with permission

Tenure-Track Faculty at Research Institutions

Incentive Structure:

- Publish papers
- Bring in grant money
- Graduate students

Timeline:

7 years to tenure (typical)

BUT actually, 6 years (so you can find a job somewhere else if you aren't getting tenure)

BUT actually, 5 years (year 6 is when your materials are evaluated)

BUT actually, 4 ½ years (since publishing takes about 6 months)

Typical grants are 3-year and 5-year grants

Funding Agencies

- National Science Foundation (NSF)
- National Institutes of Health (NIH)
- Department of Energy (DOE)
- Department of Defense (DOD)
- United States Department of Agriculture (USDA)
- Private Foundations, e.g., Andrew Mellon Foundation

Anatomy of a Proposal

Cover page, Title, PIs/Co-PIs, Project summary: one page

brief project description (executive summary) - explain to reviewers and program officer what you plan to do, why it'll work, and how it'll help.

Project Description: 15 pages on average, intro, usually 3-4 project objectives, intellectual merit, implementation plan, broader impacts, management plan, evaluating progress

Broader Impacts

- Advancement of scientific knowledge
- Activities that contribute to achievement of societally relevant outcomes
- Participation of women, persons w/ disabilities, & underrepresented minorities in STEM
- Improved STEM education and educator development at any level
- Increased public scientific literacy and public engagement with science & tech
- Improved well-being of individuals in society
- Development of a diverse, globally competitive STEM workforce
- Increased partnerships between academia, industry, and others
- Improved national security; Increased economic competitiveness of the US
- Enhanced infrastructure for research and education

Your broader impacts are judged on what you've already done

More Proposal Content

- Management Plan
- Advisory Committee
- Timeline and milestones
- Sustainability plan: What happens when the grant ends?
- Budget: people, things, subcontracts (subject to IDC), participant support
- Budget justification
- Data Management Plan
- Letters of Commitment, letters of collaboration
- Biographical sketches
- Current and Pending Support
- Conflicts of Interest

Probability of Success

National Science Foundation FY2018: 24% Overall

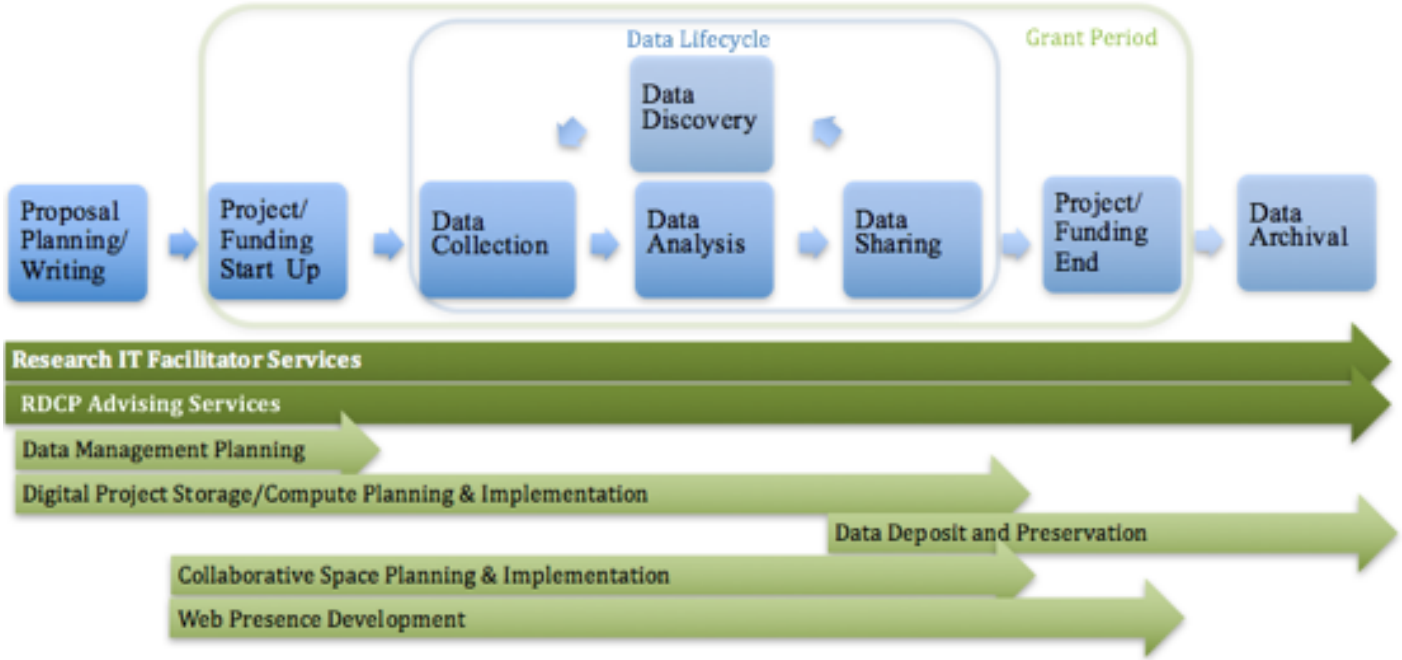
BIO 25%, CSE 23%, EHR 21%, ENG 19%,
GEO 37%, MPS 29%, SBE 23%

Funding is governed by the Law of Large Numbers:

You have to submit lots of proposals to get any funding

http://dellweb.bfa.nsf.gov/awdfr3/default.asp

An Introduction to Research Facilitation



The Research Lifecycle

Acknowledgements: Claire Mizumoto - content shared with permission

Introduction to Research Facilitation

The Main Goal: To Move Research Forward

- Researcher-facing professionals contribute to the efforts of the campus:
  - To increase funding opportunities
  - To train researchers to optimize use of tools, software, services, infrastructure
    - > Efficiency
    - > Effectiveness
  - Support collaborative efforts
  - Make way for the next generation
    - As researchers
    - As facilitation professionals

Introduction to Research Facilitation

Skills Needed to be Successful

- Learn to see things from another point of view
  - Researcher
  - Executive / Administration
  - Service Provider
- Speak the same language
- Create a communication conduit
  - Be the translator
  - Be Switzerland
  - Consult, advise
  - Be the matchmaker


Protected data
CPU GPU VPN
NSF NIH
Globus
Parallel Serial
FIONA
Storage
XSEDE OSG
Endpoint
Datasets
Cost
Compliance
Size of Data
Compute Time
IRB
Affiliates
Deadlines
Grant Management
Database
Analysis
Patient Data
Budget
Timeline
Collaborators
Script
Discipline
Publishing
Facilitator
Researcher

Introduction to Research Facilitation

Develop These Skills, Too

- Get researchers as productive as possible, as quickly as possible
- You get one hour with a standard consultation
- Delve into the researcher’s world
  - Go broad, not deep
  - Understand enough to respond with something helpful, to speak intelligently

Develop in-depth, comprehensive knowledge on a wide range of products, services, and service providers, and be able to identify appropriate solutions that will best meet the needs of the research project.

Introduction to Research Facilitation

Partnering With Facilitators on Your Campus

- Find out where the researcher-facing professionals are are on your campus
  - Often in HPC centers
  - Cyberinfrastructure focus
  - Some campuses now have areas of Research IT
- What type of activities do they provide?
- How can you partner with them?

National Efforts

  - Engaging more higher education and research institutions
  - Professionalization of research supporting positions

Cyber(Security)Ambassadors

—

Communications Skills Workshop

CyberAmbassador Project

- NSF training grant
  - Provide professional skills training to cyber infrastructure (CI) professionals
  - Focus on communication, teamwork, leadership
  - Overarching goal to support interdisciplinary research
- Collaborators
  - Tau Beta Pi - Engineering Honor Society
  - Software / Data Carpentry
  - Campus Champions (XSEDE)
  - CaRCC
  - Blue Waters
  - CIMER and National Research Mentoring Network (NRMN)
  - Research Security Operations Center NSF Grant Number 1840034
  - Local and regional universities

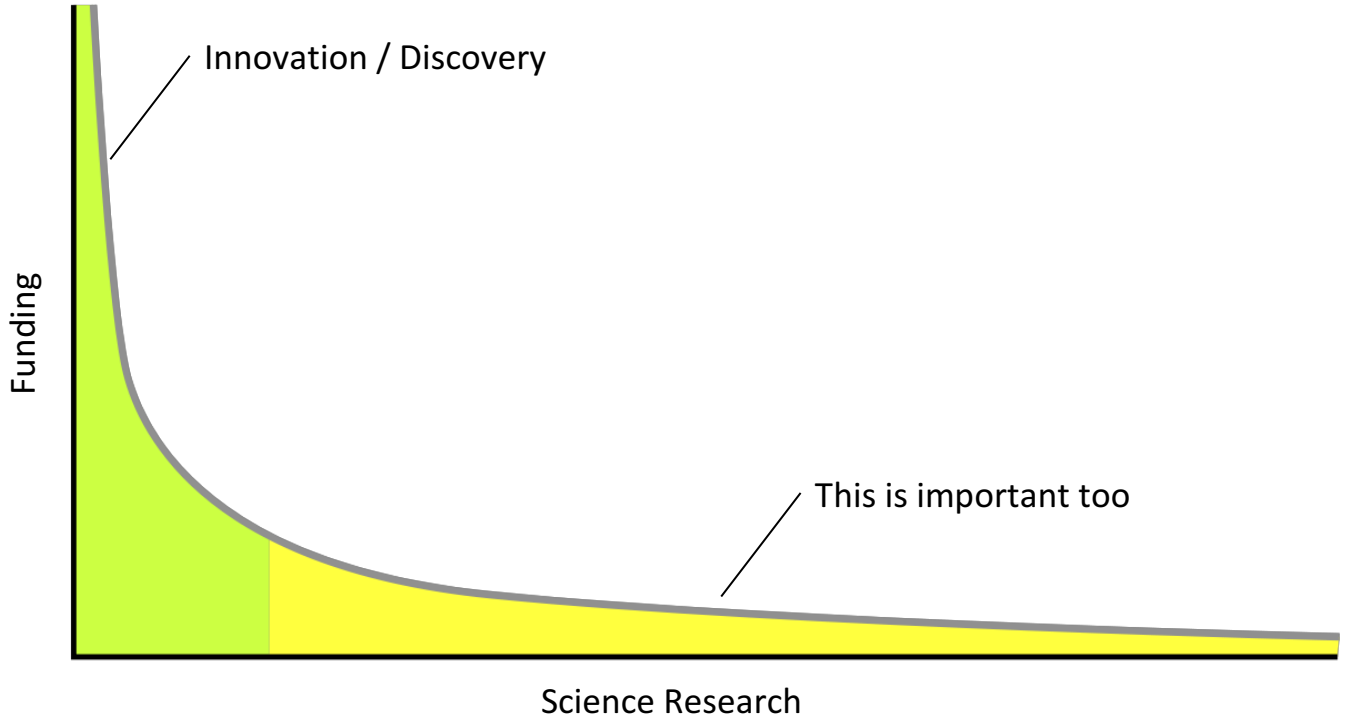
How do we support research computing?

Listen to the researchers and understand their needs

Lower barriers to entry

Support rapid innovation

Supporting both innovation and the long tail of science



Effective Communication…

…occurs when information is both shared and understood

…builds trust and fosters relationships

…helps the speaker communicate needs and goals

…helps the listener understand and participate in solutions

Improving Communication Skills

- Communication is a major topic of research
  - Effective communication skills can be taught!
  - There are tools (algorithms) that apply across many scenarios
  - Role playing / rehearsal activities are effective learning tools
  - Practice is most effective in context

First Contact: Consulting

Intake Interviews

Office Hours

Requests for Expertise

Opening an Intake Interview

Hi, my name is Kelly. I need to download and install some export controlled data for my research project. What factors do I need to consider and how do I get started?

What questions do you ask?

Why are Intake Interviews Important?

Good Intake Interviews (Research Facilitators)

- Ask about their research
- Take the time that is needed
- Identify the fundamental problem(s) being solved
- Work to describe the problem in terms of security concerns
- Avoid Jargon
- Ask how they think the problem should be solved
- Assess their ability
- Identify assumptions (yours and theirs)
- Learn to say “No” without saying No

Closing an Intake Interview

- Identify next actions
- Set a time to complete action items
- Follow up

Intake Example: Port Forwarding

The Context
Nan is a graduate student that has put together a website based on a downloaded program template to gather data from citizen scientists; the website is running in a docker instance on an office desktop. Nan has requested a hole in the university firewall in order to be able to publicize the website and start gathering research data. This data is considered “IRB Exempt” and Nan has made sure there is no information on the desktop that is sensitive.

The Speaker
Jamie is a Security Expert who works for the university’s IT Department. After a quick package review, Jamie can see that Nan’s program is running on an out of date version of node.js which could potentially open up the desktop to all types of known security bugs. Access to a desktop inside of the University Firewall could then open up additional security vulnerabilities.

Jamie's First Attempt

Jamie says: "I've reviewed your project and I think you can fix the security problems. You just need to update all of your node.js packages to their latest versions and implement a security procedure to ensure that they do not get out of date in the future. Here is a website to help you get started. Once things are patched we can start the paperwork for opening up the firewall."

- Did Jamie use any Jargon?
- How well do you think Nan now understands the problem?
- What response will Nan likely have to this explanation?

Jamie's Second Attempt

Jamie: "I've looked though your project and found some potential security concerns. The problem is that even though you do not have any sensitive data on your desktop, there is a possibility that a nefarious person could gain access to your system and once inside the university firewall cause problems on other systems. I can help you update the software. Have you considered using cloud resources instead? Do you know what cloud resources are?"
Nan: "No"
Jamie: "Since you are using docker, cloud resources will let you run your instance on a computer outside of the university. This will insulate the university from security vulnerabilities. It has the added benefit of helping you scale the software if you end up getting a lot of users."

Breakout Activity: Intake Interviews

- It's time to practice some intake interviews using role-playing. For this activity, you will work in groups and rotate through three roles:
  - **The Security Expert:** The Security Expert's goal is to successfully identify the core problem described in the scenario and identify next steps.
  - **The Actor:** has the role of communicating the problem following the guidelines given in the scenario. Have fun acting and trying to make the Security Expert work a little before you "give in" and the scenario ends.
  - **The Coach:** has the role of observing the interactions between the Research Expert and Actor and providing feedback to both participants. Coaches are free to look at all three role cards for the scenario, and are encouraged to "pause" the scenario and offer advice to keep things on track.

Example 1 - DOS Batch File

You have a distinguished professor running old code that runs a data collection instrument. It's running a DOS batch file that interacts with hardware via some ancient compiled code that can't be tracked down anymore. The system has been collecting data on the same hardware for 20 years and the professor is worried any changes will introduce changes in the consistency of the data or introduce non-reproducible data.

You set up a meeting with the professor to explain the problem. Offer alternatives for maintaining the existing environment but discuss how change is inevitable and how to deal with the reproducibility and availability issues. Explain the situation to the professor about why the older systems must be retired, and get the professor to agree to move their research to the new hardware.

Example 2 - Backup Barrage

You have set up a meeting with a senior professor who is known to be high strung (bordering on abusive) and difficult to work with. This professor's instruments produce 5-10TB per day of data when in production, but s/he recently suffered a ransomware incident, fortunately on a secondary system. The professor has been barraging the help desk with emails and phone calls complaining about the need for an immediate, free, backup solution. After much investigation you discover the professor's lab has a 10 gig connection to the campus network.

You have scheduled a meeting, so that you can let the professor know what backup options exist, that none of them are free, and that s/he will need to pay for faster network electronics.

Acknowledgements

- The CyberAmbassadors Team: Dirk Colbry, Katy Luchini Colbry, Julie Rojewski, Astri Briliyanti and TJ Nguyen
- This material is based upon work supported by the National Science Foundation under Grant No. 1730137. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Refreshment Break: 2:30 – 3:00 p.m., Foyer, Second Floor

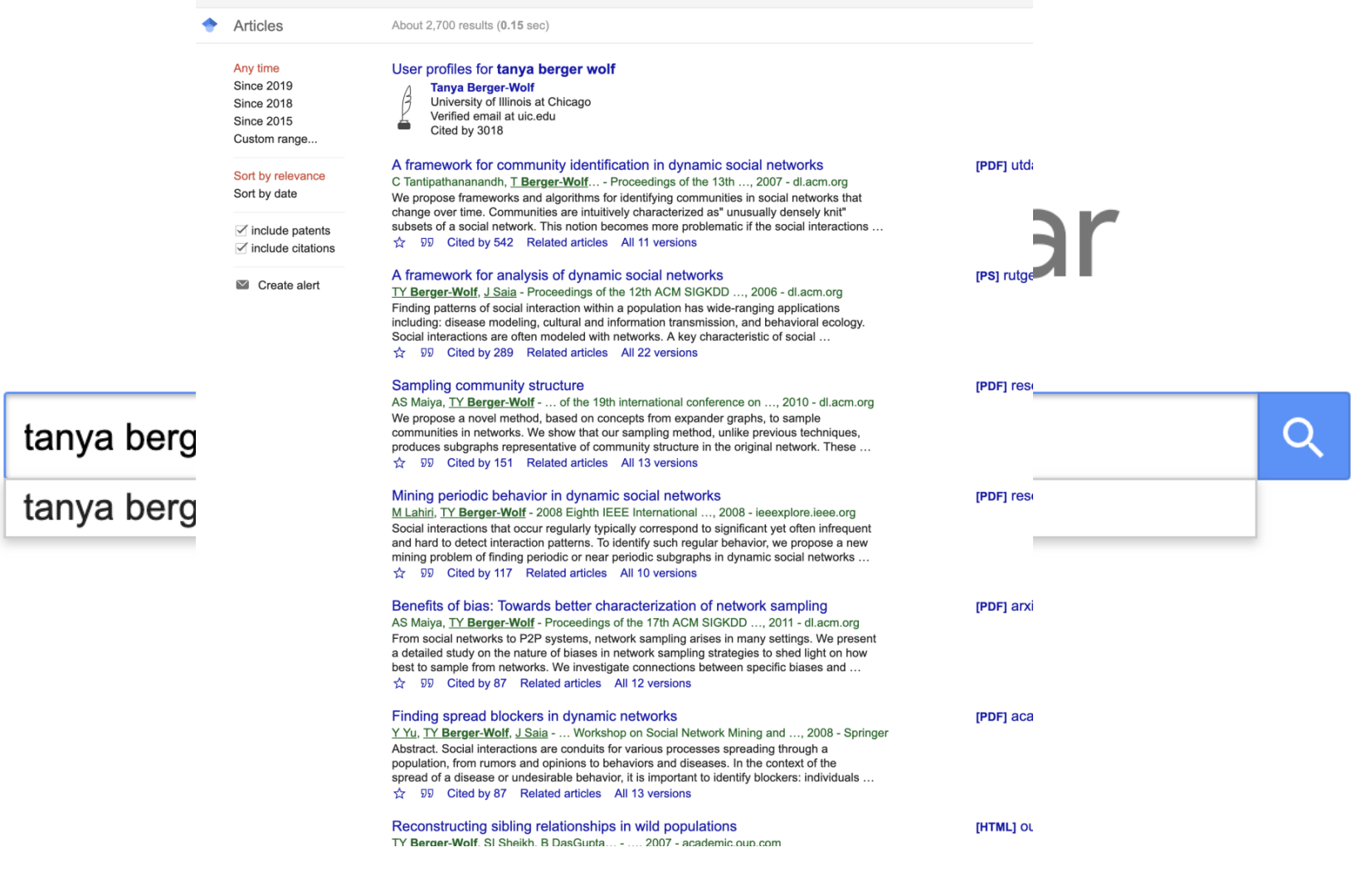
Facilitation Exercise

- Dr. Tanya Berger-Wolf, Professor, Department of Computer Science, University of Illinois at Chicago
- Area of study: *Flu pandemic, political microtargeting, behavioral response to predator presence, species genetic diversity. Populations contain intricate connections that occur on time scales ranging from milliseconds to generations. At the Laboratory for Computational Population Biology, we explore the growing interface between Population Biology and Computer Science, from genetics to social interactions.*

Pre-meeting Prep

- ❏ Google PI - watch any YouTube Interviews
- ❏ Read lab web pages
- ❏ Contact ORA/G&C for copies of relevant grants / contracts
- ❏ Find out if there's a data management plan on file, or perhaps a technology control plan, or an IRB proposal
- ❏ Talk to unit staff who may or may not support PI and lab
- ❏ Review recent publications (esp. If they reference hardware or software)
- ❏ Collaborators?

- https://www.youtube.com/watch?v=4McrCDioiCc
- https://drive.google.com/drive/folders/1WOhctiNp-szv9722qoEocbNiR3meqGrT?usp=sharing
- https://www.wildbook.org/doku.php?id=security


Articles
About 2,700 results (0.15 sec)
Any time
Since 2019
Since 2018
Since 2015
Custom range...
Sort by relevance
Sort by date
include patents
include citations
Create alert
User profiles for tanya berger wolf
Tanya Berger-Wolf
University of Illinois at Chicago
Verified email at uic.edu
Cited by 3018
A framework for community identification in dynamic social networks
C Tantipathananandh, T Berger-Wolf... - Proceedings of the 13th ..., 2007 - dl.acm.org
We propose frameworks and algorithms for identifying communities in social networks that change over time. Communities are intuitively characterized as" unusually densely knit" subsets of a social network. This notion becomes more problematic if the social interactions ...
Cited by 542 Related articles All 11 versions
[PDF] utd
A framework for analysis of dynamic social networks
TY Berger-Wolf, J Saia - Proceedings of the 12th ACM SIGKDD ..., 2006 - dl.acm.org
Finding patterns of social interaction within a population has wide-ranging applications including: disease modeling, cultural and information transmission, and behavioral ecology. Social interactions are often modeled with networks. A key characteristic of social ...
Cited by 289 Related articles All 22 versions
[PS] rutge
Sampling community structure
AS Maiya, TY Berger-Wolf - ... of the 19th international conference on ..., 2010 - dl.acm.org
We propose a novel method, based on concepts from expander graphs, to sample communities in networks. We show that our sampling method, unlike previous techniques, produces subgraphs representative of community structure in the original network. These ...
Cited by 151 Related articles All 13 versions
[PDF] res
Mining periodic behavior in dynamic social networks
M Lahiri, TY Berger-Wolf - 2008 Eighth IEEE International ..., 2008 - ieeexplore.ieee.org
Social interactions that occur regularly typically correspond to significant yet often infrequent and hard to detect interaction patterns. To identify such regular behavior, we propose a new mining problem of finding periodic or near periodic subgraphs in dynamic social networks ...
Cited by 117 Related articles All 10 versions
[PDF] res
Benefits of bias: Towards better characterization of network sampling
AS Maiya, TY Berger-Wolf - Proceedings of the 17th ACM SIGKDD ..., 2011 - dl.acm.org
From social networks to P2P systems, network sampling arises in many settings. We present a detailed study on the nature of biases in network sampling strategies to shed light on how best to sample from networks. We investigate connections between specific biases and ...
Cited by 87 Related articles All 12 versions
[PDF] arxi
Finding spread blockers in dynamic networks
Y Yu, TY Berger-Wolf, J Saia - ... Workshop on Social Network Mining and ..., 2008 - Springer
Abstract. Social interactions are conduits for various processes spreading through a population, from rumors and opinions to behaviors and diseases. In the context of the spread of a disease or undesirable behavior, it is important to identify blockers: individuals ...
Cited by 87 Related articles All 13 versions
[PDF] aca
Reconstructing sibling relationships in wild populations
TY Berger-Wolf, SI Sheikh, B DasGupta... - ... 2007 - academic.oup.com
[HTML] ou
tanya berg
tanya berg

Activity

This is our interview / facilitation portion with Dr. Berger-Wolf.

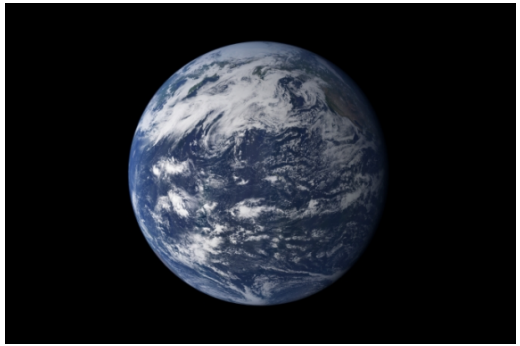
Everyone should be prepared to red-flag (er, plate) us. Each table should have a note taker, we will report out on Dr. Berger-Wolf’s requirements when finished.

Commentary - Feedback from

Workshop:

TBW:

Returning to Security World



Security tools: programs, templates, support

Terms of Art: Securing the Science DMZ

Next Steps: building a community for support

Toolkits for the Security Professional

Trusted CI:
The NSF Cybersecurity Center of Excellence

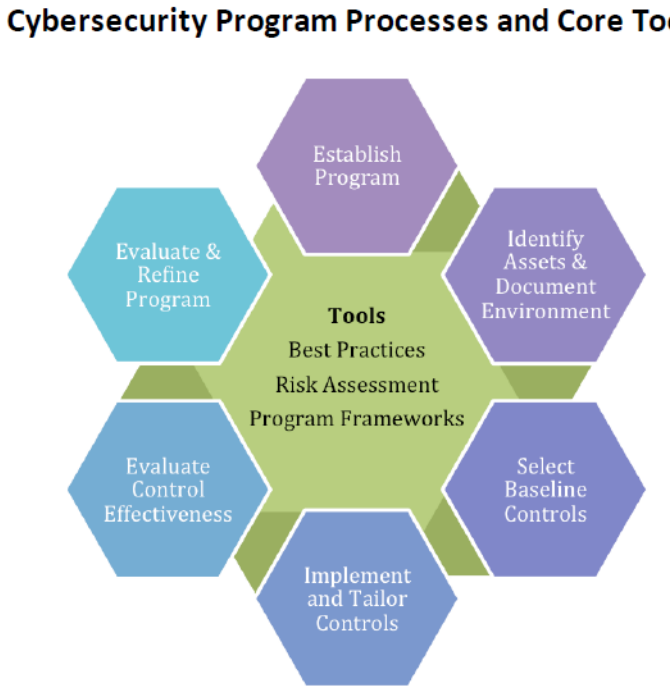
Our mission: to provide the NSF community a coherent understanding of cybersecurity's role in producing trustworthy science and the information and know-how required to achieve and maintain effective cybersecurity programs.

https://trustedci.org

TRUSTED CI
THE NSF CYBERSECURITY CENTER OF EXCELLENCE

- Trusted CI engages directly with science and research CI to solve security problems.
- Research CI includes HPC, control systems, one-of-a-kind instruments, highly distributed systems, land/air/sea-based systems, extreme environments, etc.
- We produce tools and guidance that open science / CI projects at any scale can use.
    - Cybersecurity Planning Guide
    - Software Engineering Guide
    - Information Security Practice Principles
- We host the NSF Cybersecurity Summit as well as the Large Facilities Cybersecurity Team to promote cybersecurity efforts throughout the NSF ecosystem

Guide to Developing Cybersecurity Programs for NSF Science and Engineering Projects (aka Cybersecurity Planning Guide)



- Currently in version 1.0
- Short guide for building a cybersecurity program for science and CI, plus a wealth of templates to jump-start policy creation
- Already in use at Gemini, LSST, and other Major Facilities
- Also in use with smaller projects such as OSG
- Highly customizable, plug and play just the parts your project needs
- https://trustedci.org/guide

Other Useful Items in Cybersecurity Planning Guide

Templates

Acceptable Use Policy Template
Access Control Policy Template
Asset Management Policy Template
Asset-Specific Access and Privilege Specification Template
Disaster Recovery Policy Template
Incident Response Policy and Procedures Template
Information Asset Inventory Template
Information Classification Policy Template
Information Security Training and Awareness Policy Template
Master Information Security Policy & Procedures Template
Password Policy Template
Physical Security Policy Template

Forms

Personnel Exit Checklist
Risk Assessment Table

Resources

Security Commodity IT in Scientific CI Projects: Baseline Controls and Best Practices
Trusted CI "Cyber Hygiene" Information Security Training Slide Deck
Developing Cybersecurity Programs For NSF Projects Slide Deck (NSF Security Summit 2014)

https://trustedci.org/guide

Trusted CI Software Engineering Guide

- Guide to Secure Software Engineering Practice for Science and Scientific Cyberinfrastructure
- Short guide for practices that will help achieve security goals of software used in science
- Maturity model helps you scale from the needs of small, one-off projects to major pieces of infrastructure
- Not a secure coding guide: these are the tools and software development processes that make secure coding more likely to succeed in practice
- Written for researchers and programmers
- Currently in draft stages, expected publication June 2019
- https://sweguide.trustedci.org/

Addressing Software Development At Any Scale

Level 1: Intended for one-off, non-network-connected software developed by one or two researchers for their own use. Goals: *software integrity, scientific reproducibility.*

Level 2: Intended for small software projects that must be used by more than their original author(s). Goals: *reliable software distribution, software integrity and scientific reproducibility.*

Level 3: Intended for most scientific cyberinfrastructure software. Goals: *reasonable trustworthiness, maintainability, reliable software distribution, software integrity and scientific reproducibility.*

Level 4: Intended for high-reliability scientific cyberinfrastructure software. Goals: *security and trustworthiness, maintainability, reliable software distribution, software integrity and reproducibility.*

Level 5: Intended for critical cyberinfrastructure. Goals: *highest levels of security, trustworthiness, maintainability, reliable software distribution, software integrity, reproducibility.*

The information Security Practice Principles

A risk-based, evidence-based framework

Information Security Practice Principles

Comprehensivity

- *Identify and account for all relevant systems, actors, and risks in the environment*

Opportunity

- *Take advantage of actor relationships, material resources, and strategic opportunities*

Rigor

- *Specify & enforce expected states, behaviors, & processes governing the relevant systems & actors*

Minimization

- *Minimize size, quantity, complexity of what is to be protected, & limit externally facing points of attack.*

Compartmentation

- *Isolate system elements, enable, control interactions strictly necessary for their intended purposes.*

Fault tolerance

- *Anticipate & address the potential compromise and failure of system elements & security controls.*

Proportionality

- *Tailor security strategies to the magnitude of the risks, accounting for the practical constraints imposed by the mission and the environment.*

https://cacr.iu.edu/about/principles.html

TRUSTED CI
THE NSF CYBERSECURITY CENTER OF EXCELLENCE

Other resources

Other Resources:

- More Templates
- Guides & Forms
- Compliance
- Presentations
- Webinar series
- Engagements

https://trustedci.org/help/

See also:

https://trustedci.org/situational-awareness/

https://trustedci.org/webinars/

https://trustedci.org/ctsc-email-lists/

http://blog.trustedci.org/

@TrustedCI

Terms of Art: The Universe of Research Computing

- XSEDE https://www.xsede.org/
- Open Science Grid (OSG) https://opensciencegrid.org/
- PERFSonar https://www.perfsonar.net/
- Globus https://www.globus.org/
- Condo / Hotel Model for Clusters
- Virtual Circuits https://www.internet2.edu/products-services/advanced-networking/layer-2-services/
- HPC vs. HTC
- GPU Cluster
- Data Transfer Node
- The Carpentries https://carpentries.org/ (software, data, and library carpentry workshops)
- Pacific Research Platform (PRP) http://pacificresearchplatform.org/ (and NRP https://bit.ly/2Y7017N)
- Jupyter Notebooks https://jupyter.org/
- **Science DMZ** https://fasterdata.es.net/science-dmz/

The Science DMZ: A Network Design Pattern for Data-Intensive Science

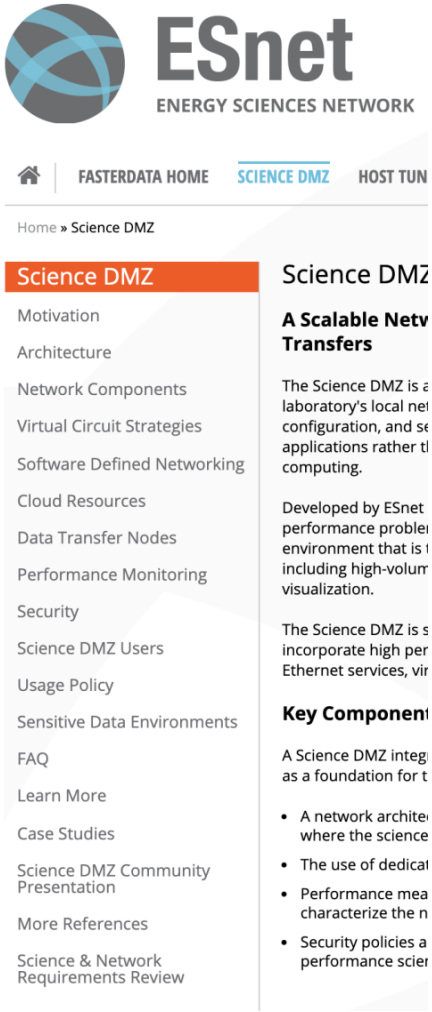
Vlad Grigorescu (ESnet)
vlad@es.net

Credits

- Eli Dart, Science Engagement @ ESnet
- Michael Sinatra, Networking @ ESnet

Tip of the Iceberg

- Some familiarity with Science DMZ as a concept
  - Not prescriptive; a design pattern and not an architecture
- How to design the Science DMZ to strengthen your security posture
- Provide you with resources
  - https://fasterdata.es.net



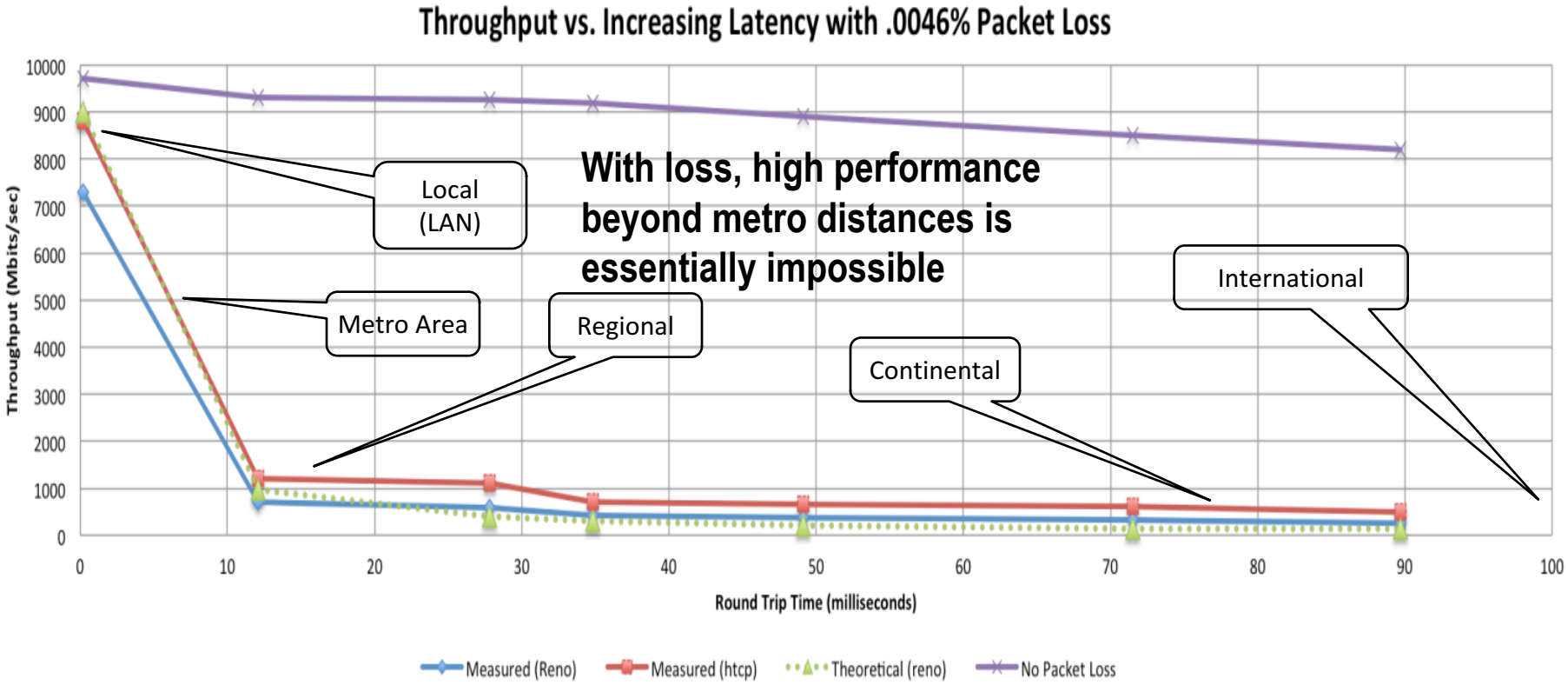
How is Data Being Transferred?

- A small number of (very) large flows
  - 10 Gigabit minimum, 100 and 400 Gigabit in production, 1 Terabit in planning stages
- GridFTP is the de-facto standard

  *“GridFTP is a **high-performance**, **secure**, **reliable** data transfer protocol optimized for **high-bandwidth wide-area networks**. The GridFTP protocol is based on FTP, the highly-popular Internet file transfer protocol.”*

- Focus on “data transfer nodes” (DTNs)
  - Systems designed from the ground up for lightning-fast disk-to-network transfers
  - https://fasterdata.es.net/science-dmz/DTN/

Effects of Packet Loss



Putting a Solution Together

- Effective support for TCP-based data transfer
  - Design for correct, consistent, high-performance operation
  - Design for ease of troubleshooting
- Easy adoption is critical
  - Large laboratories and universities have extensive IT deployments
  - Drastic change is prohibitively difficult
- Cybersecurity – defensible without compromising performance

Science DMZ Security Myth

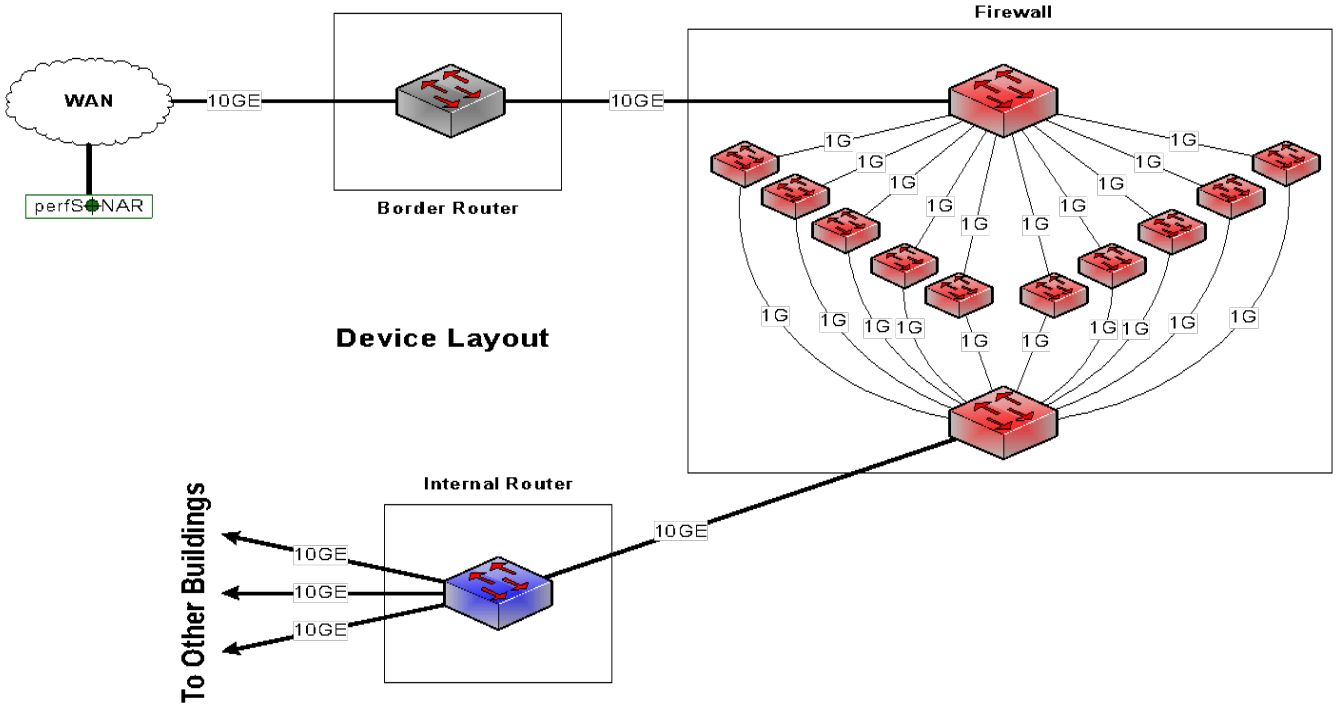
- **The big myth:** The main goal of the Science DMZ is to avoid firewalls and other security controls.
  - Leads to all sorts of odd (and wrong) claims like:
    - “Our whole backbone is a Science DMZ because there is no firewall in front of the backbone.”
    - “The Science DMZ doesn’t allow for **any** security controls.”
    - “The Science DMZ requires a default-permit policy.”

- **The reality:** The Science DMZ emphasizes reducing degrees-of-freedom, reducing the number of network devices (including middleboxes) in the path, eliminating devices that can’t perform, and ensuring that the devices that remain in the path are capable of large-scale data-transfer caliber performance.

From Myth to Reality

- Contrary to myth, the Science DMZ *is a security architecture.*
- The Science DMZ is a form of security *control,* not something to be controlled.
- At the same time, the Science DMZ enables us to do a better job of risk-based security through segmentation.
- Borrow ideas from traditional network security (Traditional DMZ)
  - Separate enclave at network perimeter (“Demilitarized Zone”)
  - Specific location for external-facing services
  - Clean separation from internal network
  - Do the same thing for science – ***Science DMZ***

How Do Firewall Appliances Work?



What is a Firewall?

Vendor Answer

- Specific appliance, with “Firewall” printed on the side
- Lots of protocol awareness, intelligence
- Application awareness
- User awareness (VPN, specific access controls, etc.)
- Designed for large concurrent user count, low per-user bandwidth (enterprise traffic)

What is a Firewall?

Security Group Answer

- “Firewall” appliance, purchased from the commercial marketplace
- The place in the network where security policy gets applied
- Owned by the security group, not by the networking group
- Primary risk mitigation mechanism

What is a Firewall?

NIST Answer (Publication 800-41 rev. 1, Sep. 2009)

“Firewalls are devices or programs that control the flow of network traffic between networks or hosts that employ differing security postures”

Problems with Firewall Appliances

- Firewalls have a lot of sophistication in an enterprise setting
  - Application layer protocol analysis (HTTP, POP, MSRPC, etc.)
  - Built-in VPN servers
  - User awareness
- Data-intensive science flows don’t match this profile
  - Common case – data on filesystem A needs to be on filesystem Z
    - Data transfer tool verifies credentials over an encrypted channel
    - Then open a socket or set of sockets, and send data until done (1TB, 10TB, 100TB, …)
  - One workflow can use 10% to 50% or more of a 10G network link
- Do we have to use a firewall?

Firewalls as Access Lists

- What does a firewall admin ask for when asked to allow data transfers?
  - IP address of your host
  - IP address of the remote host
  - Port range
  - ***That looks like an ACL to me – I can do that on the router***

- No special config for advanced protocol analysis – just address/port

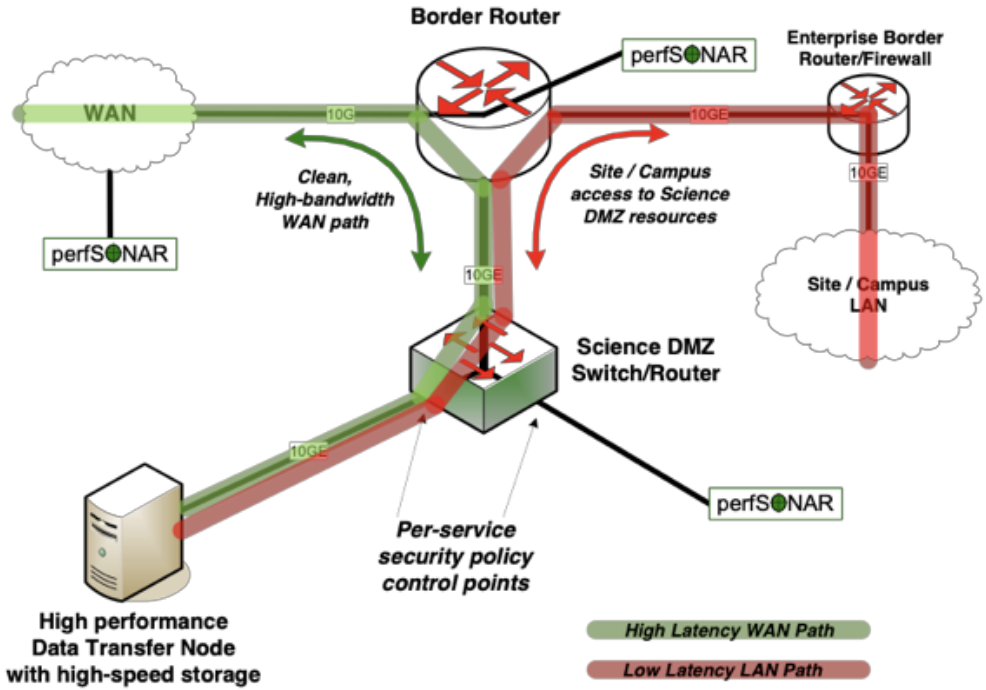
Security Without Enterprise Firewalls

- Data intensive science traffic interacts poorly with enterprise firewalls
- Does this mean we ignore security? ***NO!***
  - We **must** protect our systems
  - We need to find a way to do security that does not prevent us from getting the science done
- ***Key point – security policies and mechanisms that protect the Science DMZ should be implemented so that they do not compromise performance***

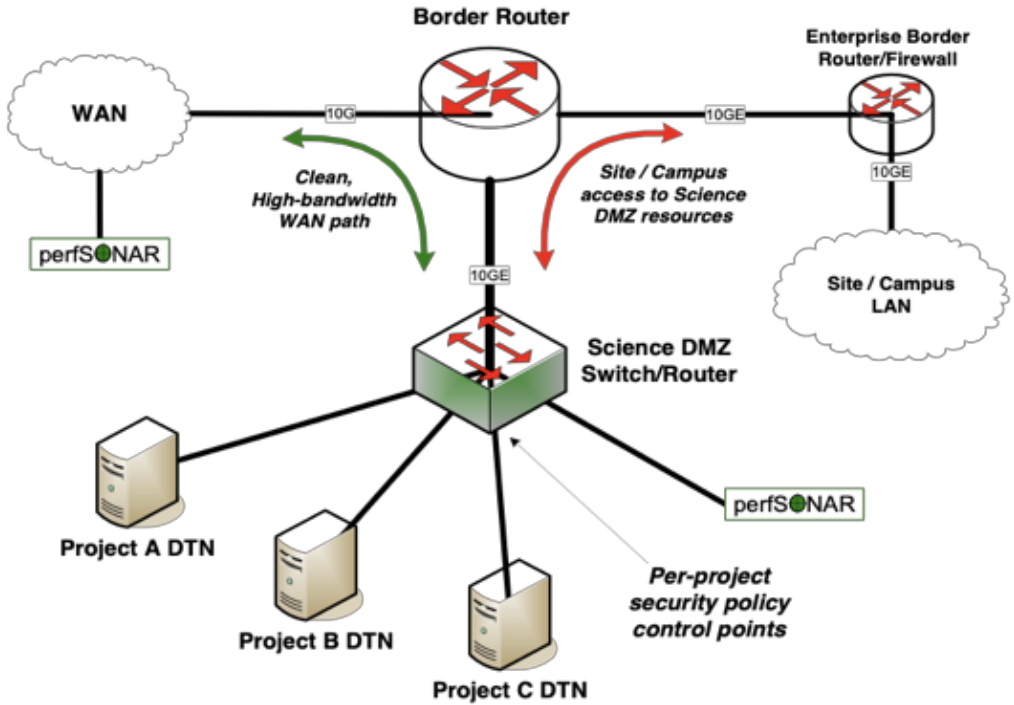
New and Emerging Firewall Designs

- Several organizations are working on ways to make firewalls better
- Some use SDN to dynamically switch approved flows around the firewall
- Some allow the firewall to control a switch directly
- Some vendors are now building firewalls to accommodate elephant flows
- ESnet hasn’t directly tested these approaches, though they look promising
- Some have significant cost

Science DMZ Example 1



Science DMZ Example 2: Multiple Projects



Other Security Mechanisms: ACLs and Applications

- **Aggressive access lists**
  - More useful with project-specific DTNs
  - Exchanging data with a small set of remote collaborators = ACL is fairly easy to manage
  - Large-scale data distribution servers = difficult/time consuming to handle (but then, the firewall ruleset for such a service would be, too)
- **Limitation of the application set**
  - Makes it easier to protect
  - Keep unnecessary applications off the DTN (and watch for them anyway using a host IDS – take violations seriously)

Other Security Mechanisms: Network Monitor

- Network Security Monitors
  - One example is Bro – https://bro.org/
  - Bro is high-performance and battle-tested
    - Bro protects several high-performance national assets
    - Bro can be scaled with clustering: https://docs.zeek.org/en/stable/cluster/
  - Other IDS/NSM solutions also available

Other Security Mechanisms: Host IDS

- Using a Host IDS is recommended for hosts in a Science DMZ
- Several open source solutions exist:
  - OSSec: http://www.ossec.net/
  - Rkhunter: http://rkhunter.sourceforge.net (rootkit detection + FIM)
  - chkrootkit: http://chkrootkit.org/
  - Logcheck: http://logcheck.org (log monitoring)
  - Fail2ban: http://www.fail2ban.org/wiki/index.php/Main_Page
  - denyhosts: http://denyhosts.sourceforge.net/

Collaboration Within the Organization

- **All stakeholders should collaborate on Science DMZ design, policy, and enforcement**
- **The security people have to be on board**
  - Political cover for security officers
  - If the deployment of a Science DMZ is going to jeopardize the job of the security officer, expect pushback
- **The Science DMZ is a strategic asset, and should be understood by the strategic thinkers in the organization**
  - Changes in security models
  - Changes in operational models
  - Enhanced ability to compete for funding
  - Increased institutional capability – greater science output

Conclusions and Implications

- Think about what the Science DMZ is trying to do.
  - Improve performance, both by removing impediments and improving the performance of the devices that must be in line
  - Apply security policies appropriate for the data and the applications being protected
  - Ease troubleshooting
  - In general, reduce degrees of freedom from science networks to increase security flexibility/options
  - Maximize performance **and** security **and** resiliency

Warning: If You Build It...

- One thing that often happens is that an early power user of the Science DMZ is the network engineering group that builds it
  - Service prototyping
  - Deployment of test applications for other user groups to demonstrate value
- The production Science DMZ is just that – production
  - Once users are on it, you can't take it down to try something new
  - Stuff that works tends to attract workload
- ***Take-home message: plan for multiple Science DMZs from the beginning – at the very least you're going to need one for yourself***

Supporting Each Other: building a community of practice

The Research Security Operations Center (ResearchSOC) has partnered with Ask.CI, the Q & A and discussion platform for all things cyberinfrastructure (CI), to create a new community and learning platform for those working to secure cyberinfrastructure crucial to open science.

To Join:

Create an account via Ask.CI: https://ask.cyberinfrastructure.org/

Find this group here: https://ask.cyberinfrastructure.org/c/rsoc

The community offers discussion, Zoom-based live learning opportunities, networking, and more. Your contributions will ensure that this becomes an active, informative community!

Thank You, Contacts, Acknowledgements

Cyd Burrows-Schilling cburrows@ucsd.edu

Michael Corn mcorn@ucsd.edu

Acknowledgements

- Dirk Colbry Michigan State University colbrydi@msu.edu
- Florence D. Hudson CACR & FDHint
- Tanya Berger-Wolf, University of Illinois, Chicago
- Claire Mizumoto, UC San Diego
- Mary Conley & Todd Stone CACR/Trusted CI
- Von Welch, CACR/Trusted CI/ResearchSOC vwelch@iu.edu